

Provider Discretion and Variation in Resource Allocation: The Case of Triage Decisions[†]

By DAVID C. CHAN AND JONATHAN GRUBER*

Productivity in health care depends on the efficient allocation of resources to tasks. In many instances, the resources are human, and the allocation mechanisms may require discretion. Primary care providers must know to whom and when to refer a patient for specialty care. Nurses must know when to notify physicians if a patient is becoming unstable. Hospital staff must decide when to discharge a patient to a skilled nursing facility instead of a home.

In this paper, we study a particularly important setting for resource allocation in health care: emergency triage. Triage—the systematic practice of prioritizing patients for care—originated on the battlefield and has become a hallmark of modern emergency care. When performing triage, a provider (often a nurse) places patients in categories of severity and influences the waiting time for medical attention. As the work flow in emergency departments (EDs) has grown in volume and complexity, and as time-sensitive treatments for life-threatening conditions have become more available, accurate and reliable systems of triage have become increasingly important to ensure that patients who require immediate attention do not suffer from delays in care.

Crucially, while systems of triage involve algorithms and define categories of patients, in practice, systems also involve the discretion of providers performing the triage on the ground. In the United States, the predominant system

of triage—the Emergency Severity Index (ESI) developed by the Agency for Healthcare Research and Quality—defines five formal levels of patient severity based on acuity and projected resource utilization (Gilboy et al. 2012). At a high level, patient assignment to ESI levels follows a simple algorithm, but this simplicity belies the complexity and detailed considerations in the 114-page official handbook. Studies of ESI assignment have also shown low interrater reliability (Hinson et al. 2019). Therefore, providers performing triage have substantial scope in allocating resources to emergency cases.

These allocation decisions may have significant implications for patient outcomes. In our data from the Veterans Health Administration (VHA), patient ED wait times vary enormously. Across the 108 EDs in our data, the average wait time for each ED varies from 60 minutes to 97 minutes between the visit-weighted twenty-fifth and seventy-fifth percentiles. Important outcomes vary widely as well. For the outcome of seven-day mortality, the visit-weighted interquartile range of ED averages is 0.3 to 0.5 percentage points, or 50 percent of the mean of 0.4 percentage points. Similarly, the probability of intensive care unit (ICU) admission varies across EDs from 0.9 to 2.2 percentage points between the weighted twenty-fifth and seventy-fifth percentiles, or 80 percent of the mean of 1.6 percentage points.

The resource-allocation intuition behind ED triage is that the potentially scarce resource of medical attention should be assigned to those who will benefit the most, and guidelines presume acuity and projected utilization should guide these decisions. Despite these implications, there is little work on these allocation decisions.

I. Empirical Approach

The VHA is the nation's largest integrated health-care system, serving nine million enrolled veterans each year. Studying detailed triage and

*Chan: Stanford University, Department of Veterans Affairs, and NBER (email: david.c.chan@stanford.edu); Gruber: Massachusetts Institute of Technology and NBER (email: gruberj@mit.edu). We thank David Card, David Silver, Peter Hull, and Chris Walters for helpful comments and suggestions. We also thank Sophie Andrews, Sydney Costantini, Johnny Huynh, Kevin Kloiber, Uyseok Lee, and Chris Lim for excellent research assistance. Chan gratefully acknowledges support from NIH DP5OD019903-01 and VA 1-I01-HX002631-01.

[†]Go to <https://doi.org/10.1257/pandp.20201032> to visit the article page for additional materials and author disclosure statement(s).

chart data, we observe 108 EDs, each corresponding to a separate VHA health-care system (or “station”). Each station (e.g., Boston, Palo Alto, Seattle) has its own potentially different ED system of resource allocation. For example, some systems may allow physicians to choose their own patients, while other systems may assign patients to physicians by a triage nurse. The common step across all systems of triage is that patients are assigned an ESI level.¹

In the VHA, key elements of the triage process are measured with standardized definitions via the Emergency Department Integration Software. In particular, we observe for each patient the front-desk registration time, the times of ESI assignment, and the time of assignment to a treating physician.² We define the time between registration and physician assignment as the patient’s *wait time*. In addition to wait time, with each ESI assignment, we observe the identity of the triage provider and the ESI level assigned. A lower ESI indicates greater severity of patient condition.³ We consider the provider associated with first ESI assignment as the triage provider of interest.

Consistent with the literature on ED triage, we are interested in measuring the quality of triage by identifying patients with acute and life-threatening illness who are nonetheless triaged as lower priority than patients without acute illness (for a review, see Hinson et al. 2019). Guidelines universally recommend that more severe patients should receive more timely care (Horwitz and Bradley 2009, Gilboy et al. 2012). We therefore focus on wait times to receive medical attention as a direct measure of resource allocation in triage. In particular, we identify “inversions” of wait-time orderings, in which sick patients receive care in a less timely fashion relative to less sick counterparts.

¹ While this discrete step is often referred to as the “triage” decision, we refer to the overall system of resource allocation prior to attention from a doctor as the triage process.

² In principle, a patient may be triaged more than once, with multiple times of ESI assignment. In the data, about 95 percent of patients are triaged only once.

³ The Gilboy et al. (2012) ESI handbook defines the following ESI levels: Level 1 requires immediate life-saving intervention. Level 2 indicates a high-risk situation, confusion, lethargy, disorientation, severe pain, distress, or many resources needed and unstable vital signs. Level 3 indicates one resource needed at unstable vital signs. Level 4 indicates stable vital signs and one resource needed. Level 5 indicates stable vital signs and no resources needed.

To construct our measure, we first identify acutely ill patients. We identify these patients by using several methods. In our first method, we use only information available *ex ante* at the time of ED presentation, including patient demographics, prior utilization, vital signs, laboratory tests, and an indicator for whether the patient arrived by ambulance, in order to predict by a random forest algorithm the probability of mortality within seven days. Importantly, we exclude diagnoses made during the ED visit in this information set. Second, we define acute illness solely on the basis of the following three life-threatening ED diagnoses—sepsis, acute myocardial infarction, and pulmonary embolism—as suggested by the medical literature (Hinson et al. 2019); we consider these diagnoses made during the ED visit as an *interim* indicator of acuity.⁴ Third, we measure acute illness by an *ex post* event, defining *ex post* acute patients as those who subsequently died within seven days or were admitted to the ICU.

After categorizing patients as acute or non-acute, we find hours when at least one acute patient and one nonacute patient arrive. If there is more than one acute patient in the hour, we select the acute patient with the shortest wait time. For comparison, if there is more than one nonacute patient in the same hour, we choose the comparator patient whose wait time is at the tenth percentile of nonacute cases in this set.

Finally, we define *inversions* as hours in which there is an acute patient who had a wait time that was longer relative to the comparator patient. Note that we can assign inversions only to *candidate* hours in which both acute and non-acute patients arrived. We assign the inversion hour to any triage provider working in that hour, weighted by the number of patients she triaged in the hour. We define a provider’s inversion rate as the percent of hours that were inversions as a proportion of the number of candidate hours.

In our data, we observe 9,348,252 ED visits with 2,563,116 patients and 11,248 triage providers. Using the *ex ante* information, we categorize the 10 percent of the visits with the

⁴ We include any diagnosis made within three days of the ED visit. We use International Classification of Diseases, Ninth Revision (ICD-9), codes [410,411] for acute myocardial infarction; ICD-9 codes [415,416] for pulmonary embolism; and ICD-9 codes 995.91, 995.92, [038,039], and 771.81 for sepsis.

highest predicted patient mortality as acute, yielding 934,808 visits with 421,411 ex ante acute patients. Alternatively, we categorize 138,427 visits (or 115,037 patients) with an interim acute ED diagnosis and 180,373 visits (or 148,385 patients) with ex post outcomes of mortality or ICU admission indicating acuity. Using the ex ante definition of acuity, we observe 612,437 candidate hours, among which 325,375 constitute hours with inversions. While the number of candidate hours differs across definitions of acuity, the overall rate of inversions remains similar across definitions, which we demonstrate below. To be sure, not all inversions necessarily constitute actual mistakes; they may sometimes reflect data entry errors. Similarly, in large EDs, some patients with very simple (and nonacute) cases may be treated quickly to improve patient flow (e.g., medication refills). We nonetheless view the overall scale of such inversions and variation in inversion rates across triage providers as important objects of discretion that have not been previously characterized.

Quasi-random arrival of patients at the ED allows us to directly compare across triage providers. We illustrate below the insensitivity of our results to controls for detailed patient characteristics. The key empirical moments of interest in this paper are (i) variation in inversion rates across triage providers and (ii) the correlations between various measures of patient acuity.

II. Results

Based on ex ante patient acuity, the mean inversion rate is 53.5 percent. The rate of inversions is also remarkably high: despite recommendations to assign wait times by patient acuity, in slightly more than half of the cases where there is a choice to be made, a nonacute patient arriving in the same hour will have a shorter wait time than the acute patient. This is consistent with smaller studies in the prior medical literature showing that more than 20 percent of patients who died after emergency care were not designated as having high acuity (Hinson et al. 2019).

We then consider variation in inversion rates across triage providers within the same station. That is, we risk adjust inversion rates by fixed effects for each station. The risk-adjusted and visit-weighted standard deviation of inversion rates is 7.7 percent. The tenth percentile of this

object is 44.1 percent, and the ninetieth percentile is 60.9 percent. Further risk adjustment for patient characteristics and time dummies does little to reduce to variation in inversion rates across providers. The visit-weighted standard deviation of fully risk-adjusted inversion rates is 6.4 percent. Furthermore, much of the variation appears to be systematic. When we restrict our analysis to 5,347 providers with at least 100 triage assignments, the visit-weighted standard deviation of fully risk-adjusted inversion rates is 6.0 percent.

In Figure 1, we show the distribution of inversion rates across providers determined using different definitions of patient acuity. The variation across providers appears similar, regardless of whether patient acuity is defined by predicted mortality from information solely available ex ante, by interim ED diagnoses, or by ex post outcomes.

To shed light on the assessments of triage providers leading to inversions, we examine ESI levels assigned to acute patients and comparator nonacute patients. Table 1 shows this joint distribution among all candidate hours when at least one acute and one nonacute patient arrive. The most acute patients receive ESI levels that do not capture their acuity (generally, ESI levels of 1 and 2 are considered appropriate for acute patients). Also surprising is that compared with their nonacute counterparts, acute patients have qualitatively similar ESI levels, only slightly indicating higher acuity. The cells on the diagonal indicate cases where both acute and nonacute patients receive the same ESI level, and the cells below the diagonal indicate cases where the acute patient receives a *less* acute (higher) ESI level than the nonacute patient. Among all candidate hours, acute patients receive the same ESI level as the comparator patient in 40.9 percent of cases and actually receive a *less* acute (higher) ESI level in 17.6 percent of cases.

Finally, we assess the relationship between inversion rates based on different measures of patient acuity, for example, using ex ante information or ex post outcomes. In Figure 2, we show that there is a strong positive correlation between inversion rates based on acute patients identified by ex ante information and rates based on acute patients identified by ex post outcomes. The strong positive relationship holds regardless of whether we perform full risk adjustment using patient characteristics.

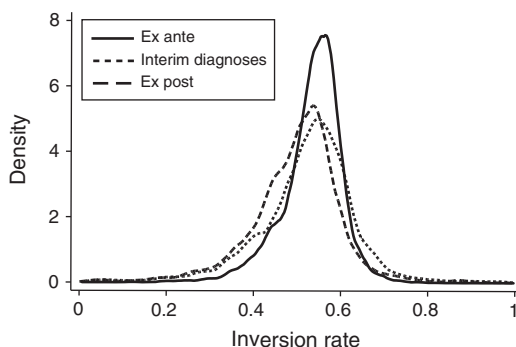


FIGURE 1. DISTRIBUTION OF INVERSION RATES

Notes: This figure shows the kernel density distribution of inversion rates across triage providers. Inversion rates are defined as the number of hours in which an inversion occurred divided by the number of candidate hours in which an inversion could have occurred, weighted by the number of visits a triage provider handled in each hour. Candidate hours are hours in which at least one acutely ill patient and one nonacutely ill patient arrive. Inversions are defined as wait-time orderings among patients arriving in the same hour, in which the shortest wait time for an acutely ill patient is longer than the tenth percentile of wait time for nonacutely ill patients. We consider three constructions of patient acuity: (i) ex ante, which uses information available prior to the ED diagnosis; (ii) interim acute ED diagnoses (sepsis, pulmonary embolism, and acute myocardial infarction) according to the medical literature; and (iii) ex post, which uses outcomes of ICU admission and mortality. Kernel densities are weighted by the number of visits per triage provider.

III. Discussion and Conclusion

There has been an enormous and exciting growth in papers in health economics focused on the determinants of patient outcomes in the medical system. Research by us and others has focused on factors such as hospital quality (Doyle et al. 2015, Hull 2018), the generosity of health insurance benefits (Chandra, Gruber, and McKnight 2010; Brot-Goldberg et al. 2017), and physician diagnostic and treatment patterns (Molitor 2018; Chan, Gentzkow, and Yu 2019). But in our view, too little attention has been paid to a key point in the health-care process: systems and behaviors within health-care delivery that determine access to care. Of inpatients, 43.8 percent arrive at the hospital through the ED, and they arrive in a frenetic and uncertain environment (Schuur and Venkatesh 2012). The process of triage can therefore be a life-or-death determination for millions of hospital patients every year.

TABLE 1—ESI LEVELS OF ACUTE VERSUS NONACUTE PATIENTS

	Nonacute ESI (%)					
	1	2	3	4	5	Total
<i>Acute ESI (%)</i>						
1	0.03	0.11	0.30	0.21	0.05	0.70
2	0.14	3.72	7.76	5.89	1.23	18.74
3	0.28	7.01	29.25	20.23	4.30	61.07
4	0.06	1.74	6.01	6.88	1.44	16.13
5	0.01	0.30	1.01	1.05	0.98	3.35
Total	0.51	12.89	44.33	34.25	8.01	100.00

Notes: This table shows the joint distribution of ESI levels for two patients arriving in the same hour: an acutely ill patient with the shortest wait time in the hour and a nonacutely ill comparator patient. We compute patient acuity by a random forest prediction using available ex ante information, such as patient demographics, prior utilization, and vital signs. We categorize acute patient visits as the 10 percent of all visits with the highest predicted patient mortality. If there is more than one acute patient in the same hour, we choose the acute patient with the shortest wait time; if there is more than one nonacute patient in the same hour, we choose the nonacute comparator patient with a wait time at the tenth percentile for nonacute patients. ESI levels are assigned by triage providers and are supposed to indicate patient acuity and expected resource utilization. Lower ESI levels indicate higher acuity or expected resource utilization. The ESI algorithm is shown in Figure 1. While this distribution is based on the ex ante definition of acuity, interim and ex post measures yield similar results.

In this paper, we provide some initial evidence on the potential importance of this triage process. Taking advantage of the quasi-random assignment of patients to ED triage nurses according to time of arrival, we examine the size and variation of triage decisions that seemingly misorder attention to patients according to their acuity. We can define these inversions on the basis of ex ante (presenting characteristics), interim (diagnosis), and ex post (outcomes) measures of patient acuity, and these measures are highly correlated.

We find surprisingly large rates of inversions counter to official triage recommendations, which specify that acutely ill patients should receive care first. Furthermore, we find large variation in inversion rates, regardless of how acute illness is defined, and consistency in triage provider inversion rates across these different measures. Surprisingly, acute patients are hardly recognized as such by assigned ESI levels. At the very least, this last finding suggests that triage providers may be misclassifying patients'

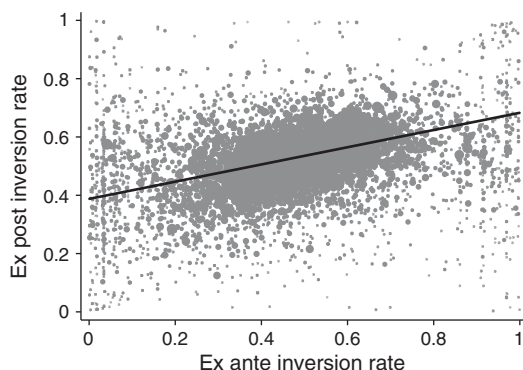


FIGURE 2. RELATIONSHIP BETWEEN INVERSION RATES

Notes: This figure shows the relationship between each triage provider's inversion rate as defined by ex post patient acuity and her inversion rate as defined by ex ante patient acuity. Each dot represents a triage provider with non-missing inversion rates based on ex ante and ex post definitions of patient acuity. This results in a sample of 7,221 triage providers with at least one candidate hour based on ex ante acuity and at least one candidate hour based on ex post acuity. The size of each dot represents the number of visits associated with each triage provider. The mean number of visits per triage provider is 4,308; the twenty-fifth and seventy-fifth percentiles are 1,692 and 5,629, respectively.

acuity. It remains to be seen whether inversions in wait time lead to worsened patient outcomes. If so, then our results point to a large scope for algorithms to reduce inversions, prioritizing patients correctly according to guidelines by forming better predictions of acuity based on patient characteristics. If not, then triage providers in some sense may “know better.” They may either form better predictions of patient acuity or recognize that, in many cases, managing a busy ED will require strategies other than simple ranking of patients on the basis of severity. In this case, algorithms that instead use the “wisdom of the crowd” could reveal effective triage strategies that are already in use (Chen et al. 2017).

We hope that these results serve as an entry point to an exciting set of questions around triage and, more broadly, resource allocation by providers and organizational systems.

REFERENCES

- Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad. 2017. “What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics.” *Quarterly Journal of Economics* 132 (3): 1261–318.
- Chan, David C. Jr., Matthew Gentzkow, and Chuan Yu. 2019. “Selection with Variation in Diagnostic Skill: Evidence from Radiologists.” NBER Working Paper 26467.
- Chandra, Amitabh, Jonathan Gruber, and Robin McKnight. 2010. “Patient Cost-Sharing and Hospitalization Offsets in the Elderly.” *American Economic Review* 100 (1): 193–213.
- Chen, Jonathan H., Mary K. Goldstein, Steven M. Asch, Lester Mackey, and Russ B. Altman. 2017. “Predicting Inpatient Clinical Order Patterns with Probabilistic Topic Models vs Conventional Order Sets.” *Journal of the American Medical Informatics Association* 24 (3): 472–80.
- Doyle, Joseph J. Jr., John A. Graves, Jonathan Gruber, and Samuel A. Kleiner. 2015. “Measuring Returns to Hospital Care: Evidence from Ambulance Referral Patterns.” *Journal of Political Economy* 123 (1): 170–214.
- Gilboy, Nicki, Paula Tanabe, Debbie Travers, and Alexander M. Rosenau. 2012. *Emergency Severity Index (ESI): A Triage Tool for Emergency Department Care*. Rockville, MD: Agency for Healthcare Research and Quality.
- Hinson, Jeremiah S., Diego A. Martinez, Stephanie Cabral, Kevin George, Madeleine Whalen, Bhakti Hansoti, and Scott Levin. 2019. “Triage Performance in Emergency Medicine: A Systematic Review.” *Annals of Emergency Medicine* 74 (1): 140–52.
- Horwitz, Leora I., and Elizabeth H. Bradley. 2009. “Percentage of US Emergency Department Patients Seen within the Recommended Triage Time: 1997 to 2006.” *Archives of Internal Medicine* 169 (2): 1857–65.
- Hull, Peter. 2018. “Estimating Hospital Quality with Quasi-experimental Data.” http://www.mit.edu/~hull/RAM_022018.pdf.
- Molitor, David. 2018. “The Evolution of Physician Practice Styles: Evidence from Cardiologist Migration.” *American Economic Journal: Economic Policy* 10 (1): 326–56.
- Schuur, Jeremiah D., and Arjun K. Venkatesh. 2012. “The Growing Role of Emergency Departments in Hospital Admissions.” *New England Journal of Medicine* 367: 391–93.